"Development of Algorithm Using Deep learning for Share Market Prices with Time Series Analysis"

Proposal submitted for the registration of Ph.D. in Computer Science.
Under the Faculty of

Submitted by

Mr. Ali Abdullah Ahmed Aldaoar
hialidaoar@gmail.com
Albidha-Yemen

Under the Guidance of
Dr. Parag Bhalchandra
(Assistant Professor)

Research center
School of Computational Sciences
S.R.T.M University Nanded, Maharashtra

Through
The Director , International Student Center

June 2019

Development of Algorithm Using Deep learning for Share Market Prices with Time Series Analysis”

INTRODUCTION :

Stock market plays an important role in nowadays financial markets. And it has attracted more and more attention in the past few decades. Many researchers from different fields pay more and more attention on it. With the development of computer science technology, some methodologies based on data mining and machine learning have been introduced to predict the stock price movement automatically. With the development of algorithmic trading and electronic trading, more and more investors noticed that the stock volatility prediction should not only focus on the prediction accuracy but also the prediction speed. The prediction accuracy is the first requirement for the designed prediction model. Higher prediction accuracy can help the market traders make a better decision. And the fast speed of real-time market transaction asks for a faster prediction speed. In the past few decades, several approaches, such as neural network and support vector machine, have been widely applied to classification and regression problems.

Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. physhological, rational and irrational behaviour, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.

Can we use machine learning as a game changer in this domain? Using features like the latest announcements about an organization, their quarterly revenue results, etc., machine learning techniques have the potential to unearth patterns and insights we didn’t see before, and these can be used to make unerringly accurate predictions.

Nowadays, social media has become a mirror that reflects people's thought and opinions to any particular event or news. , since the number of individual investors has increased , research efforts to support these people in financial markets have been actively conducted . Whats important for investors when it comes to investing is know the companies future performance

forecasts from hints such as follow .It is common to browse summaries of financial statement of companies as a means to know (predict) the future performance forecast of the company .

The stock market is one of the most important components forming a country's economy . Through IPO - Initial Public offering . a company is able to raise a substantial amount of money to expand businesses . As is well known, stock market is a very complex non- linear time series system . It is assumed that behaviors of stock market to be shown in the future could be predicted with previous information given in the history .

This follows Machine Learning is a term coined around 1960 composed of two words -machine corresponding to a computer , robot or other device , and learning an activity , or event patterns , which humans are good at. We will apply machine learning to marketing analysis as a simple use case with business application. We will survey popular machine learning approaches applying each to marketing analysis. We observe some approaches are more interpretable as measuring return on investment (ROI) of different marketing actions, and other approaches provide more accurate sales forecasts.

LITERATURE REVIEW:

Machine learning and artificial intelligence techniques are being used in conjunction with data mining to solve a plethora of real world problems. These techniques have proven to be highly effective, yielding maximum accuracy with minimal monetary investment and also saving huge amounts of time.To add to their annual income, nowadays, people have started looking at stock investments as a lucrative option. With expert guidance and intelligent planning, we can almost double our annual revenue through stock returns. That said, stock investment still remains a risky proposition for the uninitiated.

Mankar et al.[1] have proposed based on the comparative study that we performed, Support Vector Machine proved to be the most efficient and feasible model in predicting the stock price movement, in favour of the sentiments of the datas. Using Machine Learning techniques for prediction purposes is inexpensive compared to the ground survey that would have been conducted otherwise to gauge the public mood. Cloud services will enable us to collect large amount of data and also store it in real time when we will get the data directly from the REST API. Classification of datas as positive, negative and neutral gives a good overview of public mood. Other factors that are affecting the public mood will be subsequently studied and incorporated in the system.

Dang et al.[2] the topic of stock trend prediction using deep learning interests many researchers and investors because the improved prediction accuracy will probably bring enormous profit. In this paper, we introduced a stock movement prediction framework to support investors in trading stock on the market. The proposed Two-stream Gated Recurrent Unit (TGRU) overall accuracy was 66.32% which outperforms the performance of the previous model including GRU, LSTM, our model has two states of learning including backward and forward so it is able to learn more useful information, especially for text processing problem. Then a sentiment Stock2Vec embedding is created by using financial dataset and Harvard IV-4 sentiment dictionary, through the experiments, the Stock2Vec embedding proved to be more effective than the original embedding method such as Glove and Word2Vec because it takes the sentiment value of the word into consideration. Moreover, the financial indicators, which are one of the most important factors in the financial analysis, are also added. The model shows its robustness in both the S&P 500 index and individual stock trend prediction. In addition, a simulation system was conducted

to compute the actual profits investors earn when they use our system; it proved to be robust against market volatility and the ability to adjust itself to the risk from the real market. This framework can also be integrated into an automated system to support investors in trading specific stocks.There are several ways to extend this work in the future. In our study, only the daily period was considered, so it would be better if the model can be applied to intraday trading which means the system could predict the trend minutes or hours after the news is published. In addition, only three financial indicators are used in this study, so exploring other indicators such as P/E ratio and earnings growth is necessary. Another weakness of the system is in the TGRU network because the complexity of the network double that of GRU so the framework requires long training time and huge computational resources.

Lin Cheng et al. [3] have proposed to produce trading strategies, we apply deep learning to predict stock price movements. We train models with historical price data and technical indicators, after which we use the prediction result to decide on a trading strategy.

Chen et al. [4] have proposed to predict stock market behavior is challenging. Nonlinear relationship among transaction data and unpredictable factors in market fluctuations make predictions difficult. Deep learning is a machine learning method suitable for solving nonlinear approximations that has been successfully applied in many fields.To achieve better stock market predictive performance, we have used a deep architecture-based model. We use oneminute high frequency transaction data from the CSI 300 futures contract (IF1704) in the Chinese stock market and carry out an empirical analysis. To show the effect of sample volume on network training and predicting, we divide the sample into three scale datasets and compare the stock price prediction of deep learning with three traditional artificial neural networks (BP, ELM, RBF). We compare their predictive fitting degree and directional predictive accuracy and find that the predictive performance of deep learning is superior to that of BP, ELM and RBF. We also find that increasing sample volume significantly increases the predictive performance of deep learning. We thus find that sample volume strongly affects stock prediction, and that deep learning performs well when applied to large data. Furthermore, deep learning does not need prior predictive information to extract features from large datasets, and this increases its usefulness in predicting stock market behavior. This result provides additional evidence that DL is an effective method of predicting stock price. Our deep learning prediction model expands our ability to analyze financial market behavior. Because there are complex relationships among

stock futures prices and such factors as the economy, politics, the environment, and culture, future research could apply complex network theory to key input variables that influence stock prices and returns. That would allow the construction of a deep learning network that would facilitate better predictive performance.

Wang et al. [5] have proposed a new learning model named MKL-ELM, which combines the market prices and the market news to predict the market price volatility. And
we compared it with some traditional methods, such as BPNN and SVM, and the B-ELM and K-ELM. We conduct the experiments on the one whole year stock market tick data of Hong Kong Exchange. Results have shown that, under the consideration of both prediction accuracy and prediction speed, MKL-ELM outperforms than the the others in most cases. And further more, since this MKL-ELM can be regarded as a multiple data source integration framework, other data sources (such as trading volume) can also be filled into this framework and help achieve much better prediction results. But if the data sources become more complicated, MKL-ELM
will become more time-consuming. In our future work, we will try to improve the efficiency of MKL-ELM with more data sources.

Junli, et al.[6] have proposed to investigated the modeling and forecasting of the fuzzy neural network (ANFIS), which is a class of adaptive networks and functionally equivalent to a fuzzy inference system. The experiment data set applied in this study consists of 244 comprehensive index values selected from Shanghai Stock Market during the period of March 15, 2004 to March 15, 2005. The total relative fitting error is less than one percent and the total forecasting error less than five percent. In addition, the ratio of correct tendency is higher than ninety percent. Based on the experiment results, the suggested fuzzy neural network could be an efficient system of forecasting financial time series.

Cheng et al.[7] have proposed to supported by National Nature Science Foundation of China [grant number 61103125]; the Doctoral Fund of Ministry of Education of China [grant number 20100141120046]; the Natural Science Foundation of Hubei Province of China [grant number 2010CDB08504]; the 111 Programme of Introducing Talents of Discipline to Universities [grant number B07037] and Wuhan University Academic Development Plan for Scholars after 1970s .

Devi et al.[8] have proposed to the data has been collected from NSE.com. The historical data for the period of five year since 2007 to 2011 were taken in to account for analysis. The BoxJenkins methodlogy is used to identify the model. The AICBIC test criteria is applied agianst the data represented in the previous to select the best the model. The best model equation is derived for all indices. The MAPE, PMAD and % Error accruacy is applied to determine the discrimination between the actual historical data and the forecast data. This papers inferences a new investment decisions or guidelines based on the minimum error percentage obtained through the above said performamce measurues .The futre forecast of each index for next few years also highlighted in this paper. It is hoped that more innovative approaches will be conducted to bring the hidden information about the stock market.

AYDIN et al.[9] have proposed to performance measurements of support vector machine and deep learning algorithms based on up-to-date CICIDS2017 dataset were presented comparatively. Results show that the deep learning algorithm performed significantly better results than SVM. We are going to use not only port scan attempts but also other attack types with machine learning and deep learning algorithms, apache hadoop and spark technologies together based on this dataset in the future.

Yalcin et al.[10] have proposed to it has been aimed to develop a deep network, recognizing both fonts and letters in Turkish. With this aim, a pre-trained network has been trained with nearly 13 thousands images. The letter recognition training accuracy, was %100, and font recognition training accuracy was %73.44, because of the similarity of the fonts. But in order to increase the font recognition percentage, a probability calculation has been used, after the network output has been found. Even if the first test image font accuracy is 14/26, because of the probability is bigger than 0.5, it has been accepted as Arial. So with this way, the recognition performance has been increased a bit more. The network has been tested with 12 images. These images contains all letters. According to the results, letter recognition with this network has nearly %100 percentage, but the accuracy of font recognition is low, as it can bee seen from Table IV. But using the probability, font recognition percentage has been increased. In future studies, a GUI will be developed and the number of tests will be increased. And the possible fonts will be shown with a probability respectively.

Abbas et al.[11] have proposed to Deep learning is an important field in machine learning using Artificial Intelligence. The using of these systems depends on its behavior and performance. The main objective of this paper is to improve the performance of this type using a combined version from Random Forest and HTM Cortical Learning Algorithm. The proposed algorithm is called RFHTMC. The results depict that the proposed algorithm can reduce the mean absolute percentage error by half. And increases the overlap duty cycle by 15%.The proposed set of rules can increase the overall performance of the deep gaining knowledge of system by means of the preceding values. And it within the identical time can maintain the values of active duty cycle. Also, it maintains the mixed integer representation of input data units and the durability of the system.

Zhong al.[12] have proposed to propose a novel deep learning architecture by stacking feature learning methods. We apply the feature learning method MFA to this framework and name it MDA.In this case, MDA can be initialized by a supervised pre-training method. Furthermore, some deep learning techniques, such as back propagation, denosing and dropout operation, are employed on MDA to improve its performance. Extensive experiments demonstrate that on data sets with limited amount of data, the performance of MDA is better than both shallow feature learning models and relevant deep learning models. Experiments on the CIFAR-10 data set show that MDA can be used in CNNs for the supervised initialization of their fully-connected layers. In the future work, we intend to exploit some other feature learning methods for the deep architecture construction and explore the different structures of this novel deep learning framework.

Zhao al.[13] have proposed to an effective time series clustering framework was proposed with application to clusterLandsat8 time series. CD-DTW distance was chosen as the similarity measure since it can eliminate the effect of the sharp noises of time series sequences. We compare the EUC-DTW and CD-DTW in the land cover classification of Landsat8 time series and the CD-DTW algorithm performed better. In the future, CD-DTW can be tested for more data in other domains and compared to more variants of DTW.

Julien al.[14] have proposed to designed a fair benchmark for quantitative assessment of time series reconstruction technique validation and comparison, which we applied for demonstration to the IDR technique. However, LTDR-V4 cloud flag has been shown to be insufficient to estimate correctly cloud-free and cloudy time series statistics. Therefore, the quantitative assessment of IDR errors should be used as indicative of expected errors, and not as a definitive validation. In a near future, we will develop an alternative approach to flag pixels as cloud-free or cloudy, based only on parameter values. Then, we plan to use the CloudSim model to compare quantitatively the most widely used methods for time series reconstruction, in order to provide a fair intercomparison of methods, as well as to provide users with a clear map of areas where given methods are fully reliable.

Lourenco al.[15] have proposed to In this work, we have described a procedure to obtain time series models for generating synthetic series of solar irradiance for the city of Petrolina in Northeast Brazil from satellite data. The obtained models were presented alongside with their validation. The validation process was composed of two steps: checking the mean square error between the satellite data histogram and the histogram from 2000 synthetic scenarios. Results shown that from the four model structures obtained in this work (AR, ARI, ARMA and ARIMA) the ARI structure is the most suitable for generating synthetic time series for this city. The synthetic scenarios generated are stochastic data that prevents the use of historic averages for PV farm planning. The use of satellite data is proven to generate satisfactory time series model for generating synthetic scenarios with averages close to the original dataset.

All previous literature reviews which mentioned interested of process to fetch data from several sources and we do to process there data by using machine learning and deep learning in order to predictive on stock market .

OBJECTIVES OF THE STUDY:

The main objective of this research is to solve a plethora of real world problems, yielding maximum accuracy with minimal monetary investment and also saving huge amounts of time to add to their annual income, such a prediction would greatly help a potential stock investor in taking informed decisions which would directly contribute to his profits.

- To provide an overview of the data sciences and related predictive work done in this area.
- To explore different techniques already used for adapting machine learning ,deep learning and time series in this area.
- To implement a technique for Machine Learning , Deep Learning and Time Series using several models.
- To analysis and performance evaluation of the implemented models.
- Analyze the models in its various lag to know which one model is giving best results in the developing economy like India where the volatility in the market is very high and market working widely on a concept of heard behavior. In later stage we will also the forecasting accuracy of Neural Network using various train function to know that which one function is giving more accurate results. In last we will also make the comparison between best model of to know that which one provide more accurate results.

PROPOSED METHODOLOGY:

There are many factors that may have impacts on the stock price volatility, such as historical prices, news about big mergers, bankruptcy of some companies and economic turmoil. The market news and stock market historical prices are widely concerned by both academic researchers and market practitioners. In order to make these data ready for the prediction tasks, some work including the data preprocessing and the data alignment must be done at first.

The time series forecasting methods [6] generally classified into two groups based on statistical concepts and computational intelligence techniques such as neural networks (NN) or genetic algorithms (GA). Statistical time series forecasting methods are subdivided as:

- Exponential smoothing methods
- Regression methods
- Autoregressive integrated moving average (ARIMA) methods
- Threshold methods
- Generalized autoregressive conditionally heteroskedastic (GARCH) methods.
- Write on deep learning in stock market.

Data Collection

Share market data from NSE , BSE and other indices will be collected .

Data Preparation

- Transform the data to stabilize the attributes.
- Find the difference if it is not stationary; successively difference
- Series to attain stationary

SIGNIFICANCE OF STUDY:

We seek to predict the stock market prices of various companies by performing sentiment analysis of NIFTY Midcap 50 Index and the social media data such as datas related to the respective companies. Stock price volatility prediction is regarded as one of the most attractive and meaningful research issues in financial market. Some existing researches have pointed out that both the prediction accuracy and the prediction speed are the most important facts in the process of stock prediction.

PROPOSED WORK PLAN:

Year-wise plan of work and targets to be achieved

Year 1:

- Literature Survey.
- Theoretical analysis of the existing techniques is applied to improve the analytic in stock market.
- To evaluate their efficiency considering the analysis.

Year 2:

- Design and implementation technique to improve the analytic in stock market using ML , DL and TSA .
- Preprocessing of the data analysis in the stock market.
- Analytics and Performance evaluation of experimental results in terms of recognition rate and prediction time.

Year 3:

- Result Analysis and Interpretation of the experimental result.
- Preparation of final synopsis and thesis.
- Submission of synopsis and thesis.

REFERENCES:

[1] Mankar, T., Hotchandani, T., Madhwani, M., Chidrawar, A., & Lifna, C. S. (2018, January). Stock Market Prediction based on Social Sentiments using Machine Learning. In *2018 International Conference on Smart City and Emerging Technology (ICSCET)* (pp. 1-3). IEEE.

[2] Minh, D. L., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access*, *6*, 55392-55404.

[3] Cheng, L. C., Huang, Y. H., & Wu, M. E. (2018, December). Applied attention-based LSTM neural networks in stock prediction. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 4716-4718). IEEE.

[4] Minh, D. L., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep learnin approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access*, *6*, 55392-55404.

[5] Wang, F., Zhao, Z., Li, X., Yu, F., & Zhang, H. (2014, July). Stock volatility prediction u multi-kernel learning based extreme learning machine. In *2014 International Joint Confere Neural Networks (IJCNN)* (pp. 3078-3085). IEEE.

[6] Li, R. J., & Xiong, Z. B. (2005, August). Forecasting stock market with fuzzy neural netv In *2005 International Conference on Machine Learning and Cybernetics* (Vol. 6, pp. 3475- IEEE.

[8] Devi, B. U., Sundar, D., & Alli, P. (2013). An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process*, *3*(1), 65.

[9] Aksu, D., & Aydin, M. A. (2018, December). Detecting Port Scan Attempts with

Comparative Analysis of Deep Learning and Support Vector Machine Algorithms. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)* (pp. 77-80). IEEE.

[10] Sevik, A., Erdogmus, P., & Yalein, E. (2018, December). Font and Turkish Letter Recognition in Images with Deep Learning. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)* (pp. 61-64). IEEE.

[11] Abbas, M. A. (2018, March). Improving deep learning performance using random forest HTM cortical learning algorithm. In *2018 First International Workshop on Deep and Representation Learning (IWDRL)* (pp. 13-18). IEEE.

[12] Zhong, G., Zhang, K., Wei, H., Zheng, Y., & Dong, J. (2019). Marginal Deep Architecture: Stacking Feature Learning Modules to Build Deep Learning Models. *IEEE Access*, *7*, 30220-30233.

[13] Zhao, Y., Lin, L., Lu, W., & Meng, Y. (2016, July). Landsat time series clustering under modified Dynamic Time Warping. In *2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA)* (pp. 62-66). IEEE.

[14] Julien, Y., & Sobrino, J. A. (2015, July). CloudSim: a fair benchmark for comparison of methods for times series reconstruction from cloud and atmospheric contamination. In *2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)* (pp. 1-4). IEEE.

[15] Lourenço, L. F. N., de Camargo Salles, M. B., Gemignani, M. M. F., Gouvea, M. R., & Kagan, N. (2017, November). Time Series modelling for solar irradiance estimation in northeast Brazil. In *2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)* (pp. 401-405). IEEE.

Chicago

Research Guide

Research Scholar

Dr. Parag Bhalchandra

Abdullah Ahmed Aldaoar